Model Archive Summary for Suspended-Sediment Concentration at Station 11455165; Miner Slough At Hwy 84 Bridge, Ca

This model archive summary summarizes the suspended-sediment concentration (SSC) model developed to compute 15-minute SSC during July 16, 2008 through May 26, 2015. The is the first suspended-sediment model developed for the site. The methods used follow U.S. Geological Survey (USGS) guidance as referenced in relevant Office of Surface Water/Office of Water Quality Technical Memorandum 2016.07/2016.10 and USGS Techniques and Methods, book 3, chap C4 (U.S. Geological Survey, 2016, Rasmussen and others, 2009). This model archive summary is in accordance with Attachment A of Office of Water Quality Technical Memorandum 2015.01 (U.S. Geological Survey, 2014).

Site and Model Information

Site number: 11455165
Site name: Miner Slough At Highway 84 Bridge, California
Location: Lat 38°17'30", long 121°37'51" referenced to North American Datum of 1983, Solano County, CA, Hydrologic Unit 18020163.
Equipment: A YSI 6-series multi-parameter water-quality sonde equipped with a model 6136 turbidity sensor was installed on July 16, 2008 and was discontinued on May 27, 2015.

Model number: 11455165.SSC.WY08.1
Model calibration data period: September 03, 2008 to March 27, 2015.
Model application date: July 16, 2008 to May 26, 2015.
Computed by: Tara Morgan-King, USGS, Sacramento, CA (tamorgan@usgs.gov)
Reviewed by: Anna Conlen, USGS, Sacramento, Ca (aconlen@usgs.gov)

Physical Sampling and Sediment Data

All sediment data were collected using U.S. Geological Survey (USGS) protocols and are stored in the National Water Information System (NWIS) database https://waterdata.usgs.gov/nwis. Discrete, boat-based samples are collected seasonally (roughly 6-12 times throughout the year) spanning the range of conditions and specifically during large sediment transport events.

Sample collection was consistent with approved field methods described in Edwards and Glysson (1999) and U.S. Geological Survey (2006). The equal-discharge-increment (EDI) method was used to determine the locations of five sampling verticals along the transect where discharge weighted suspended-sediment samples were collected. Each sampling vertical is located at the centroid of increments representing 20% of the total flow (5 verticals). Due to the tidal nature of the site, the EDI method was used to collect discharge-weighted samples because velocities were not always isokinetic (based on Table 4-5 from TWRI09A4, U.S. Geological Survey, 2006). A boat-based discharge measurement was collected immediately before sampling to determine the location of each sampling vertical. A Federal Interagency Sediment Project US D-96 bag sampler with a Teflon nozzle and plastic bag was used to collect depth-integrated samples at each vertical across the cross section. The average sampling depth was roughly 17 feet. Velocities during the model calibration data period ranged from -0.88 ft/s to +3.5 ft/s. Sediment at this station was mostly fines (93% fines on average) and any potential sampling bias was considered minimal. Sampling bias can occur with the presence of sand. Percent fines were analyzed starting in 2013 and ranged from 84% to 100% and most samples were greater than 90% fines.

Samples were analyzed for SSC and sand/fines split at the USGS-Sediment Analysis Laboratory Santa Cruz, California. Each of the 5 verticals were analyzed individually at the lab for quality control purposes. Samples were typically not composited due to rapidly changing, tidal conditions. The average SSC from the 5 depth-integrated verticals was computed and used in the calibration dataset. When the SSC at a vertical was deemed erroneous, a manual average was computed from fewer than 5 verticals and noted in the database. On 3/12/2009, 2/23/2012, and 12/5/2013 the averages were computed from 4 verticals. Additionally, EDI sets from the same day are typically not considered replicates. Same day samples were considered unique unless the total sample collection time was less than 45 minutes for both sets. Sediment results are publicly available online.

All sediment data were reviewed and approved in the USGS Water-Quality System database.

Surrogate Data

Continuous 15-minute turbidity data, reported in formazin nephelometric turbidity units (FNU), 15-minute discharge, and/or hourly tidally filtered discharge were evaluated as possible explanatory variable for SSC. Turbidity data logged from July 16, 2008 through May 27, 2015. The operational limit for the 6136 sensor is 1000 FNU and was never exceeded throughout the deployment. The surrogate turbidity data were computed, reviewed, and approved before using in the sediment calibration model following Wagner and others (2006) and U.S. Geological Survey (2010). The published timeseries data are publicly available online.

Model Calibration Dataset

The approved time-series data spanning the dates of the sediment constituent dataset were retrieved from NWIS-TS (Rasmussen and others, 2009). The USGS Surrogate Analysis Index Developer Tool (SAID) was used to pair the concurrent continuous data with the discrete SSC data (Domanski and others, 2015). Concurrent turbidity and discharge data were selected for each discrete SSC sample by selecting the closest value within +/- 15 minutes.

A total of 38 SSC samples, representative of the cross section, were compiled. The final model calibration dataset consists of only 37 concurrent surrogate measurements and SSC samples due to missing turbidity data on December 5, 2014. Data were collected from September 3, 2008 to March 27, 2015 throughout the range of site conditions. Summary statistics and complete model calibration dataset are provided in the following sections.

Model Development

Simple linear regression (SLR) models and multiple linear regression (MLR) models were assessed using methods described in Helsel and others (2020). A variety of models were evaluated: Model 1) linear model with one explanatory variable (turbidity), Model 2) $\log_{10}$ transformed model with one explanatory variable (turbidity), Model 3) repeated medians method (Helsel and Hirsh, 2002) using one explanatory variable (turbidity), Model 4) linear model with two explanatory variables (turbidity and discharge), and Model 5) $\log_{10}$ transformed model with two explanatory variables (turbidity and discharge).

Diagnostic statistics and plots for model review were computed using a combination of Matlab, SAID, and the R environment software (Matlab, 2019, Domanski and others, 2015, R Core Team, 2018) to determine the best model. Table 3 in Rasmussen and others (2009) shows the best statistical diagnostics to help evaluate regression models. The best fit model was chosen based on residual plots, model standard error, coefficient of determination (R^2), adjusted R^2_a (a measure for comparing models with differing numbers of explanatory variables because it is adjusted for the degrees of freedom), root-mean-squared error (RMSE), mean square prediction error (MSPE), significance tests (p-values), and prediction error sum of squares (PRESS) statistics.

Values for the diagnostic statistics were computed for the five models and are included in the table below. RMSE and PRESS statistics cannot be used to compare regressions with different response variable units, so R^2, R^2_a, MSPE values, and residual plots were used as the main determinants of model strength when comparing $\log_{10}$-transformed and untransformed models. Models 1-3 included turbidity as a surrogate and models 4 and 5 include both turbidity and discharge. The $\log_{10}$-transformed model residual plots were more homoscedastic (constant variance) and the probability plots were more normally distributed compared to the other models. While the $\log_{10}$-transformed models had the better normal distribution, the inclusion of discharge as a second explanatory variable was not statistically significant in a multi-log model because the p-value of the streamflow coefficient was >0.05. It was determined that discharge was not necessarily correlated at this site due to tidal influence and sediment conditions in the watershed.

No.	Model Formula	R^2	R^2_a	RMSE	PRESS	MSPE	n	p-value Q	Model Type
Model 1	SSC ~ TURB	0.96	0.96	20.32	18992	24.15	37	na	SLR linear
Model 2	$\log_{10}$SSC ~ $\log_{10}$TURB	0.95	0.95	0.13	0.68	30.30	37	na	SLR $\log_{10}$
Model 3	SSC ~ TURB	0.92	0.92	27.57	47411	32.77	37	na	repeated median
Model 4	SSC ~ TURB + Q	0.98	0.98	14.60	8443	17.4	37	0.00	MLR linear
Model 5	$\log_{10}$SSC ~ $\log_{10}$TURB + $\log_{10}$Q	0.95	0.95	0.13	0.74	30.03	33	0.57	MLR $\log_{10}$

Residual and probability plots were evaluated to verify the best model and the simple $\log_{10}$-transformed model (Model 2) was selected. The RMSE and MSPE statistics for the log model were computed from the re-transformed variables with better results than the linear model (RMSE 17.4, PRESS, 12153, and MSPE 20.7).

Flagged observations from the SAID outlier test criteria were evaluated. Studentized residuals from the model were inspected for values greater than 3 or less than negative 3. Values outside of that range are considered potential extreme outliers. The studentized residuals were reviewed and none of the samples were deemed extreme outliers.

Model Summary

The final model for suspended-sediment concentration at site 11455165 is a simple log_{10}-transformed regression model based on 37 concurrent measurements of turbidity and cross-sectional SSC samples collected over 7 years. The SLR model is shown below with basic model information, regression coefficients, correlation, summary statistics and Duan's bias correction factor (Duan, 1983):

Linear Regression Model	Coefficient of Determination (R^2)
$log_{10}SSC = 0.435 + 0.87 * log_{10}Turb$	0.95

where

SSC is suspended-sediment concentration, in milligrams per liter, and
Turb is turbidity, in formazin nephelometric units, measured with a YSI 6136.

Because SSC was transformed during regression model development, the computed prediction may be biased and needs to be multiplied by a non-parametric smearing bias correction factor (BCF) which is shown below.

Model	Start date	End date	Linear Regression Model	*BCF*
1	7/16/2008	5/26/2015	$SSC = 10^{0.435} \times Turb^{0.87} \times BCF$	1.04

Parameter (continuous time-series)	Minimum	Maximum
Turbidity (FNU) entire record	0	554
Computed SSC (mg/L)	0	*692/397

*Extrapolation defined as computation beyond the range of the model calibration dataset may be used to extrapolate no more than 10 percent outside the range of the sample data used to fit the model and is therefore limited. The extrapolation threshold for this model is 397 mg/L.

Suspended-Sediment Concentration Record

The SSC record is computed using this regression model in the USGS National Real-Time Water Quality (NRTWQ) Web site. The complete record can be found at http://nrtwq.usgs.gov/ca.

Model

logSSC = + 0.87 * logTURB + 0.435

Variable Summary Statistics

	logSSC	SSC	logTURB	TURB
Minimum	0.602	4.0	0.114	1.3
1st Quartile	1.080	12.0	0.820	6.6
Median	1.640	44.0	1.520	33.0
Mean	1.600	84.1	1.340	54.4
3rd Quartile	2.080	120.0	1.850	70.7
Maximum	2.560	361.0	2.460	286.0

Basic Model Statistics

Number of Observations	37
Standard error (RMSE)	0.129
Average Model standard percentage error (MSPE)	30.2
Coefficient of determination (R^2)	0.953
Adjusted Coefficient of Determination (Adj. R^2)	0.952
Bias Correction Factor (BCF)	1.04

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(>|t|)
(Intercept)	0.435	0.0486	8.94	1.47e-10
log.TURB	0.870	0.0327	26.60	8.12e-25

Correlation Matrix

	Intercept	E.vars
Intercept	1.000	-0.899
E.vars	-0.899	1.000

Outlier Test Criteria

Leverage	Cook's D	DFFITS
0.162	0.194	0.465

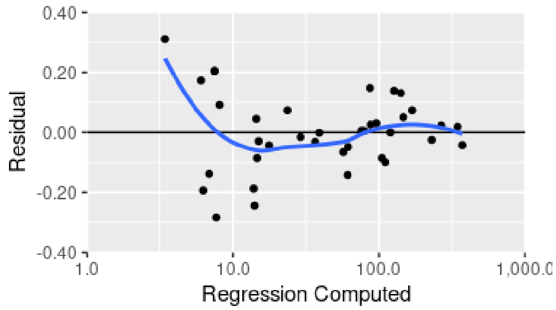
Flagged Observations

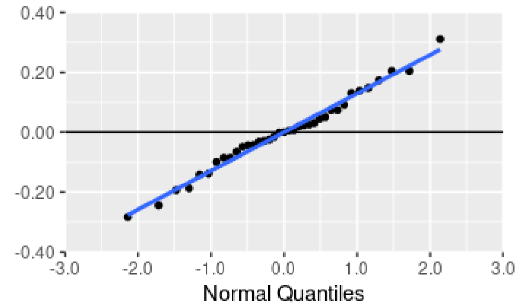
Date	Time	logSSC	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
10/9/2008	13:06	0.602	0.796	-0.194	-1.56	-1.60	0.0815	0.108	-0.476
12/3/2008	13:16	0.602	0.886	-0.284	-2.27	-2.43	0.0700	0.195	-0.666
12/5/2013	9:07	1.080	0.874	0.205	1.65	1.69	0.0714	0.104	0.468
9/11/2014	11:27	1.080	0.874	0.205	1.65	1.69	0.0714	0.104	0.468
3/27/2015	9:09	0.845	0.534	0.311	2.57	2.81	0.1230	0.462	1.050

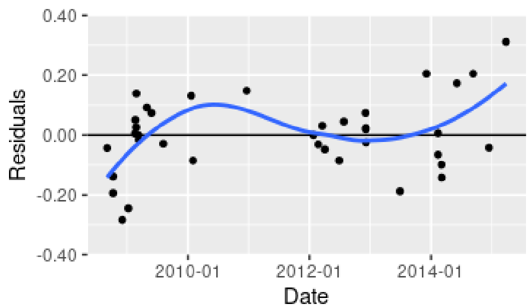
<u>Plots of log_{10}SSC and explanatory variables and residual diagnostic plots</u>

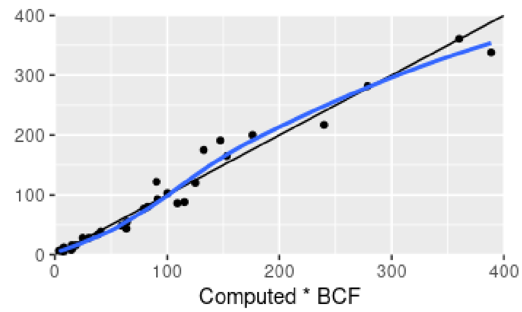
Plots were generated using the model archive summary application developed by Patrick Eslick of the USGS Kansas Water Science Center.

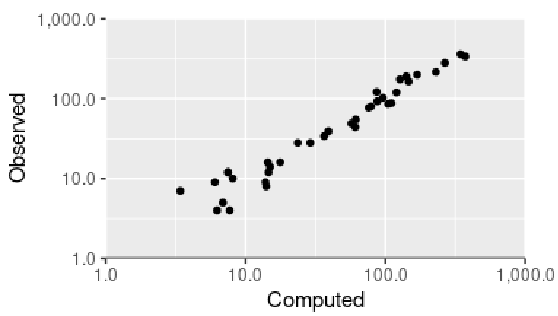
Statistical Plots

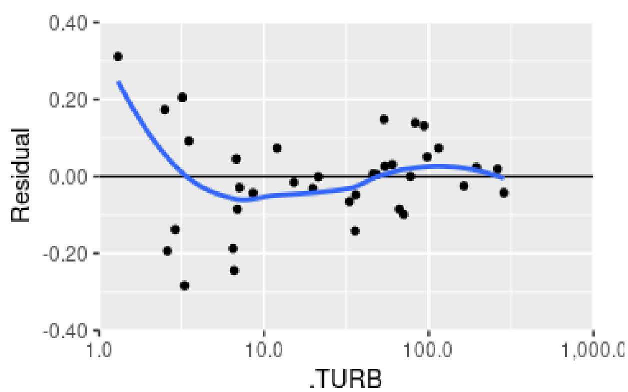


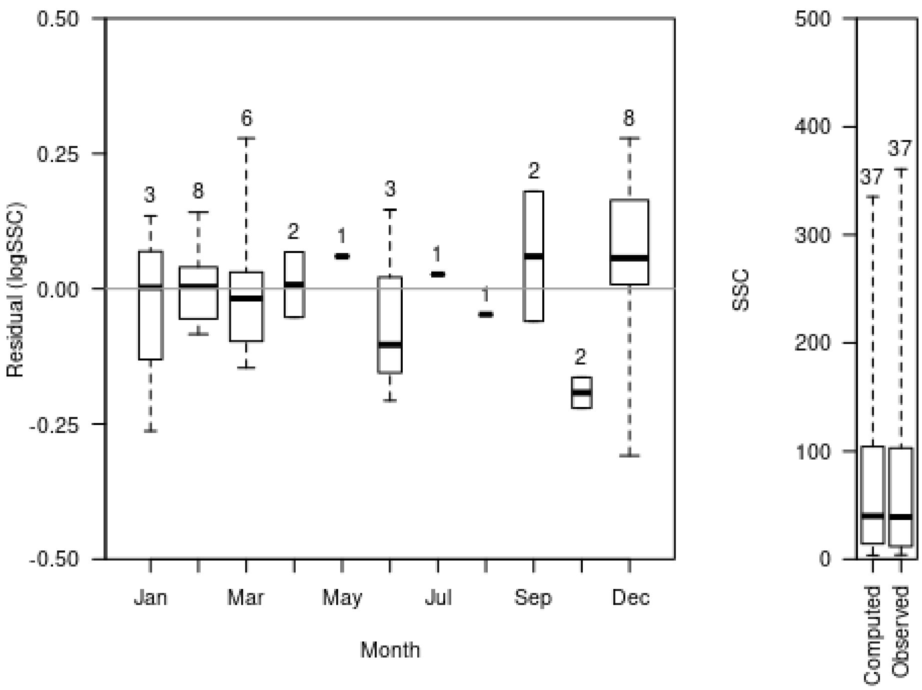


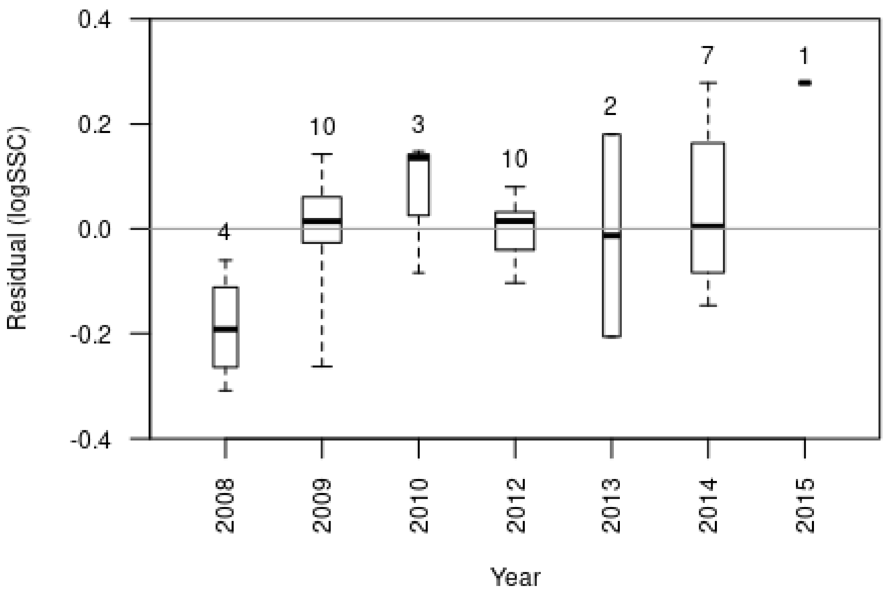




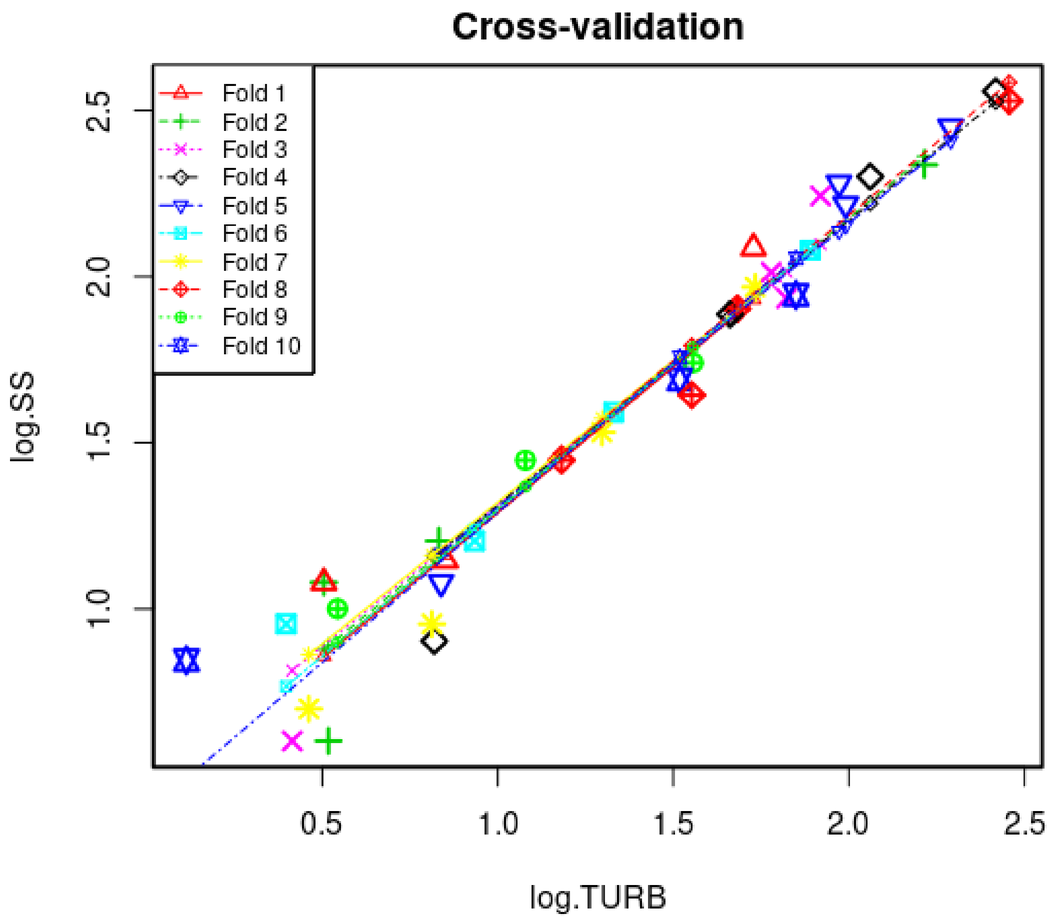




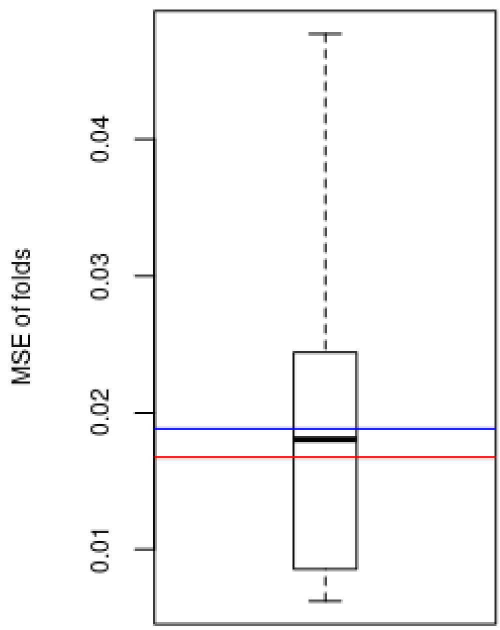

Residual (logSSC)
0.50
0.25
0.00
-0.25
-0.50
3
8
6
2
1
3
1
1
2
2
8
Jan
Mar
May
Jul
Sep
Dec
Month
SSC
500
400
300
200
100
0
37
37
Computed
Observed


Residual (logSSC)
0.4
0.2
0.0
-0.2
-0.4
4
10
3
10
2
7
1
2008
2009
2010
2012
2013
2014
2015
Year

Cross Validation




```
               Minimum MSE of folds:  0.00621
                  Mean MSE of folds:  0.01880
                Median MSE of folds:  0.01800
               Maximum MSE of folds:  0.04770
(Mean MSE of folds) / (Model MSE):  1.13000
```




```
Red line - Model MSE

Blue line - Mean MSE of folds
```

Model-Calibration Data Set

Obs #	Date/Time	Log_{10} SSC	Log_{10} TURB	SSC	TURB	Computed Log_{10} SSC	Computed SSC	Residual	Normal Quantile	Censored Values	% fines
1	9/3/2008 12:00	1.20412	0.93450	16	8.6	1.24760	18	-0.04348	-0.49068	--	--
2	10/9/2008 12:00	0.69897	0.46240	5	2.9	0.83695	7	-0.13798	-1.03413	--	--
3	10/9/2008 13:06	0.60206	0.41497	4	2.6	0.79570	7	-0.19364	-1.47659	--	--
4	12/3/2008 13:16	0.60206	0.51851	4	3.3	0.88576	8	-0.28370	-2.14120	--	--
5	1/9/2009 13:38	0.90309	0.81954	8	6.6	1.14761	15	-0.24452	-1.71677	--	--
6	2/18/2009 9:46	1.90309	1.68215	80	48.1	1.89794	82	0.00515	0.06743	--	--
7	2/19/2009 13:38	2.21748	1.99123	165	98	2.16679	153	0.05070	0.56821	--	--
8	2/24/2009 11:42	1.96848	1.73320	93	54.1	1.94234	91	0.02614	0.34355	--	--
9	2/25/2009 11:29	2.24304	1.91908	175	83	2.10403	133	0.13901	1.03413	--	--
10	3/11/2009 8:59	1.59106	1.33041	39	21.4	1.59199	41	-0.00092	-0.06743	--	--
11	3/12/2009 8:25	1.44716	1.18184	28	15.2	1.46275	30	-0.01559	-0.13518	--	--
12	4/30/2009 11:05	1.00000	0.54407	10	3.5	0.90799	8	0.09201	0.82636	--	--
13	5/27/2009 12:36	1.44716	1.07918	28	12	1.37345	25	0.07371	0.64932	--	--
14	8/7/2009 11:25	1.14613	0.85126	14	7.1	1.17519	16	-0.02907	-0.27288	--	--
15	1/21/2010 14:07	2.28103	1.97174	191	93.7	2.14984	147	0.13120	0.92525	--	--
16	2/1/2010 13:12	1.93450	1.82217	86	66.4	2.01973	109	-0.08524	-0.82636	--	--
17	12/20/2010 15:00	2.08636	1.72835	122	53.5	1.93813	90	0.14823	1.15690	--	--
18	1/25/2012 12:06	2.07918	1.89042	120	77.7	2.07910	125	0.00008	0.00000	--	--
19	2/23/2012 12:14	1.53148	1.29667	34	19.8	1.56263	38	-0.03115	-0.34355	--	--
20	3/19/2012 11:41	2.01284	1.77887	103	60.1	1.98208	100	0.03076	0.41599	--	--
21	4/3/2012 12:01	1.74036	1.55630	55	36	1.78847	64	-0.04811	-0.56821	--	--
22	6/29/2012 12:02	1.07918	0.83885	12	6.9	1.16440	15	-0.08522	-0.73496	--	--
23	7/26/2012 8:43	1.20412	0.83251	16	6.8	1.15889	15	0.04523	0.49068	--	--
24	12/3/2012 14:40	2.30103	2.06070	200	115	2.22722	176	0.07381	0.73496	--	--
25	12/5/2012 9:45	2.45025	2.29003	282	195	2.42671	279	0.02354	0.27288	--	--
26	12/5/2012 14:17	2.55751	2.41830	361	262	2.53828	360	0.01923	0.20354	--	--
27	12/6/2012 12:31	2.33646	2.21484	217	164	2.36130	240	-0.02484	-0.20354	--	--
28	6/28/2013 8:45	0.95424	0.81291	9	6.5	1.14184	14	-0.18760	-1.30015	--	94.4
29	12/5/2013 9:07	1.07918	0.50515	12	3.2	0.87414	8	0.20505	1.47659	--	--
30	2/12/2014 10:08	1.88649	1.66181	77	45.9	1.88025	79	0.00624	0.13518	--	92.7
31	2/12/2014 14:32	1.69020	1.51851	49	33	1.75560	59	-0.06541	-0.64932	--	92.1
32	3/5/2014 12:19	1.94448	1.84942	88	70.7	2.04344	115	-0.09896	-0.92525	--	98.9
33	3/6/2014 10:24	1.64345	1.55267	44	35.7	1.78531	64	-0.14186	-1.15690	--	98.4
34	6/5/2014 9:17	0.95424	0.39794	9	2.5	0.78088	6	0.17336	1.30015	--	92.9
35	9/11/2014 11:27	1.07918	0.50515	12	3.2	0.87414	8	0.20505	1.71677	--	88.4
36	12/15/2014 9:21	2.52892	2.45637	338	286	2.57139	389	-0.04247	-0.41599	--	94.2
37	3/27/2015 9:09	0.84510	0.11394	7	1.3	0.53385	4	0.31125	2.14120	--	90.5

USGS Parameter and Method Code Definitions

Model Inputs:
SSC: Suspended sediment concentration (SSC) in mg/L (80154-SED10, 80154-SED16)
Turb: Turbidity in FNU (63680-TS087)
Model Outputs:
SSC: Suspended sediment concentration by regression (99409-SED15)

MAS App Version 1.0

References

Domanski, M.M., Straub, T.D., and Landers, M.N., 2015, Surrogate Analysis and Index Developer (SAID) tool (version 1.0, September 2015): U.S. Geological Survey Open-File Report 2015–1177, 38 p., https://pubs.usgs.gov/of/2015/1177/ofr20151177.pdf.

Duan, N., 1983, Smearing estimate-A nonparametric retransformation method: Journal of the American Statistical Assocation, v. 78, no. 383, p. 605-610.

Edwards, T.K. and Glysson, G.D., 1999, Field methods for measurement of fluvial sediment: U.S. Geological Survey Techniques of Water-Resources Investigations, book 3, chap. C2, 89 p., Available from: https://pubs.usgs.gov/twri/twri3-c2/pdf/TWRI_3-C2.pdf.

Helsel, D.R., Hirsch, R.M., Ryberg, K.R., Archfield, S.A., and Gilroy, E.J., 2020, Statistical methods in water resources: U.S. Geological Survey Techniques and Methods, book 4, chap. A3, 458 p., https://doi.org/10.3133/tm4a3. [Supersedes USGS Techniques of Water-Resources Investigations, book 4, chap. A3, version 1.1.]

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources-Hydrologic analysis and interpretation: U.S. Geological Survey Techniques of Water-Resources investigations, book 4, chap. A3, 510 p.

Levesque, V.A., and Oberg, K.A., 2012, Computing discharge using the index velocity method: U.S. Geological Survey Techniques and Methods 3-A23, 148 p. Available from: https://pubs.usgs.gov/tm/3a23/.

MATLAB, 2019, version 9.6.0 (R2019a), Natick, Massachusetts: The MathWorks, Inc.

R Core Team, 2018, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, Available from: https://www.R-project.org/.

Rasmussen, P.P., Gray J.R., Glysson G.D., Ziegler A.C., 2009. Guidelines and procedures for computing time-series suspended-sediment concentrations and loads from in-stream turbidity-sensor and streamflow data: U.S. Geological Survey Techniques and Methods book 3, chap. C4, 53 p., Available from: https://pubs.usgs.gov/tm/tm3c4/pdf/TM3C4.pdf.

U.S. Geological Survey, 2006, Collection of water samples (ver. 2.0): U.S. Geological Survey Techniques of Water-Resources Investigations. book 9, chap. A4. Available from: https://pubs.usgs.gov/twri/twri9a4/twri9a4_Chap4_v2.pdf.

U.S. Geological Survey, 2014, Policy and guidelines for archival of surface-water, groundwater, and water-quality model applications: Office of Groundwater Technical Memorandum 2015.02, Office of Surface Water Technical Memorandum 2015.01, Office of Water Quality Technical Memorandum 2015.01, Available from: https://water.usgs.gov/admin/memo/SW/sw2015.01.pdf

U.S. Geological Survey, 2016, Policy and guidance for approval of surrogate regression models for computation of time series suspendedsediment concentrations and loads: Office of Surface Water Technical Memorandum 2016.07, Office of Water Quality Technical Memorandum 2016.10, Available from: https://water.usgs.gov/admin/memo/QW/qw2016.10.pdf.

Wagner RJ, Boulger RW, Jr, Oblinger CJ, Smith BA. 2006. Guidelines and standard procedures for continuous waterquality monitors: station operation, record computation, and data reporting: U.S. Geological Survey Techniques and Methods 1-D3. Available from: https://pubs.usgs.gov/tm/2006/tm1D3/pdf/TM1D3.pdf.